DOCUMENT RESUME

ED 464 914 TM 033 842

AUTHOR Hertz, Norman R.; Chinn, Roberta N.
TITLE The Role of Deliberation Style in Standard Setting for Licensing and Certification Examinations.
PUB DATE 2002-04-00
NOTE 28p.; Paper presented at the Annual Meeting of the National Council on Measurement in Education (New Orleans, LA, April 2-4, 2002).
PUB TYPE Reports - Research (143) -- Speeches/Meeting Papers (150)
EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Decision Making; Evaluation Methods; *Evaluators; *Interaction; Interrater Reliability; *Licensing Examinations (Professions); *Standard Setting; Standards
IDENTIFIERS *Rater Effects

ABSTRACT

Nearly all of the research on standard setting focuses on different standard setting methods rather than the interaction of group members and the instructions given to group members. This study explored the effect of deliberation style and the requirement to reach consensus on the passing score, on rater satisfaction, and on postdecision acceptance in three different standard setting procedures. Forty-seven licensed psychologists served as raters. Three methods were used: (1) percentage with prior group ratings (Method A); (2) yes-no consensus with prior group ratings (Method B); and (3) yes-no without prior group ratings (Method C). Results for Method C produced a passing score that was viewed as acceptable and fair and easy to understand and implement. Implications of the results are discussed in terms of social influences on group processes. (Contains 23 references.) (Author/SLD)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 464 914

Running Head: ROLE OF DELIBERATION STYLE IN STANDARD SETTING

The Role of Deliberation Style in Standard Setting

For Licensing and Certification Examinations

Norman R. Hertz and Roberta N. Chinn

HZ Assessments

Folsom, California

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

R. Chinn

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

1

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

✓ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

TM033842

BEST COPY AVAILABLE

Presented at the annual meeting of the National Council on Measurement in Education, New Orleans, Louisiana, April 2002.

Abstract

Nearly all of the research on standard setting focuses on different standard setting methods rather than the interaction of group members and the instructions given to group members. The present study explored the effect of deliberation style and the requirement to reach consensus on the passing score, on rater satisfaction, and on post-decision acceptance in three different standard setting procedures. Three methods were used: Method A - percentage with prior group ratings, Method B - yes-no consensus with prior group ratings, and Method C - yes-no without prior group ratings. Results showed that Method C produced a passing score that was viewed as acceptable and fair, and easy to understand and implement. Implications of the results are discussed in terms of social influences on group processes.

The underlying assumptions of standard setting continue to be a particular concern for regulatory agencies because the passing score on a licensing or certification examination is the ultimate criteria for determining an individual's eligibility for licensure or certification. While a number of standard setting methodologies have weathered psychometric challenges, it is not uncommon for regulatory boards to question the recommended passing score and subsequently request confidence intervals around the passing score so that they can select a score that is politically palatable to their stakeholders. Psychometricians oblige because they are unable to defend with certainty that there is an absolutely correct passing score.

The standard setting literature reports in numerous studies that it is rare for different standard setting methods to result in comparable passing scores (Jaeger, 1989). Even within variations of the same standard setting method (Angoff, 1971), some have reported similar passing scores (e. g. Impara & Plake, 1997, Plake, Impara, & Irwin, 2000) and others have not (e. g., Behuniak, Archambault, & Gable, 1982, Fehrmann, Woehr, & Arthur, 1991, Chinn & Hertz, 2002). Since many standard setting methods entail social interactions amongst the raters who set the passing score, perhaps a relevant factor that should be explored is the influence of group discussion and exposure to others' positions on decision that the members of a group make.

Social influences could be responsible for the inconsistencies in standard setting studies. Both Fitzpatrick (1989) and Miller (1989) note that when members of a group initially are disposed to favor one side of an issue or the other, in the course of discussion, they will arrive at a consensus opinion position that is more extreme in the favored direction than the average of their prediscussion opinion positions. Regardless

of whether or not a consensus opinion position is required, the private postdiscussion opinion position of the group is, on the average, more extreme that the group's prediscussion opinion position (Fitzpatrick, 1989, p. 316).

Others have examined the influences of the cognitive value of information shared amongst group members during decision making (Wittenbaum, Hubbell, & Zuckerman, 1999), group norms as a moderating factor in the value of information shared in the group (Postmes, Spears, Cihangir, 2001, Winquest & Larson, 1998), group agendas (Stasson, Kameda, & Davis, 1997), sociocognitive networks (Kameda, Ohtsubo, & Takezawa, 1997), and resolution strategy (Clark, Anand, & Roberson, 2000). There may be also be pressures of uniformity that hinder individuals' raters ability to learn during group interaction thereby producing small "group-to-individual transfer" of knowledge and skills for deriving solutions in evidence-driven groups (Kameda & Sugimori, 1993).

Concept of Deliberation Style

In discussing the role of social influences on standard setting, it may be useful to broaden our understanding of group processes and focus on other decision-making contexts. We reasoned that group decision-making processes in jury deliberation are not all that different than those traditionally considered in high-stakes examinations (e.g., academic advancement, employment, and licensing). In many cases, the consequences of erroneous judgments are equally severe.

All standard setting methods currently require some sort of deliberation to arrive at a decision by expert raters. With the Angoff standard setting method, as it is typically employed, raters form individual preferences before discussing their preferences as a

group. The mock jury studies of Davis, Kameda and others (Davis, Holt, Spitzer, & Stasser, 1981, Davis, Kameda, Parks, Stasson, & Zimmerman, 1989, Kameda, 1991, Kameda, and Sugimori, 1995) have similar group processes that could be particularly relevant to the group processes in standard setting. Deliberation style is not necessarily limited to exchanges of individual preferences (Kameda, Ohtsubo, & Takezawa, 1997) but could focus on group process that precedes deliberation. In an evidence-driven jury, deliberation begins with an exchange of views of facts rather than individual preferences. By contrast, a verdict-driven jury allows each juror to form individual preferences before the initiation of deliberation.

The influence of deliberation style can be significant in a standard setting context because there is no "correct" passing score, especially on a licensing or certification examination. Davis et al. (1981) found that minorities in juries which used a verdict driven rule were more likely to change their personal opinions than minorities in juries which used an evidence-driven rule. In a subsequent study, Davis et al. (1989) found that group consensus processes are fairly resistant to procedural issues, such as conformity pressures. Davis' findings could be explained in terms of the effect of a single vs. two-step procedures. Single-stage procedures, such as the verdict-driven rule, may have prevented local majorities from exerting influence over local minorities (Kameda and Sugimori, 1995).

Kameda (1991) found that groups who arrived at a verdict after exchanging their views at the start of deliberation arrived at a higher standard (lower guilty rate) than those groups who arrived at a verdict after exchanging preferences before deliberation. Kameda notes that members in the minority faction, who have no solid basis for their

positions, will have difficulty defending their positions against pressure from the majority faction. Moreover, when the decision alternatives are dichotomous, it may be impossible for expert raters to satisfy both factions at the same time. The most likely result is that the minority faction yields to the majority faction.

Kameda's findings are particularly relevant to standard setting studies because the verdicts derived from an evidence-driven process were perceived as being fairer and allowed more opportunities for minority opinion members to participate in group discussion than verdicts derived from a verdict-driven process.

Literature Related to Group Decision Making

Some studies focused on group decision-making processes per se suggest additional dynamics may be involved in group process. Postmes, et al. (2001) investigated the impact of group norms for maintaining consensus versus norms for critical thought on the quality of group decisions. They described group norm as a standard or rule that is accepted by members of the group as applied to themselves and other group members, prescribing appropriate thought and behavior within the group. Their overall finding was there was a paradoxical effect of group norms. They reported that cohesiveness is associated with normative pressures to conform and the drive for consensus implies intolerance toward dissent and intellectual independence of group members. They also reported that the content of group norms may make group members express their allegiance to the group not only by conforming and striving for consensus, but sometimes through independent and critical thought. In summary, their finding was that the content of the group norms was the critical factor that affected the quality of the group's decision.

Winquist and Larson (1998) designed a study to explore the effects of shared versus unshared information on the quality of group decisions. Shared information was defined as information that every group member held; unshared information was defined as information that only a few held. Only when groups pooled their unshared information did the information influence the quality of the group decision. The more unshared information the groups pooled the more likely the group would make the better decision. They identified two mediating variables as the prediscussion choice preferences and information pooling during discussion. Their model posited that the impact of shared information on group decision making is mediated primarily by member's prediscussion preferences while the impact of unshared information is mediated primarily by group discussion.

Winquist and Larson's findings suggest that the influence of group discussion should be strongest when members do not have uniform prediscussion preferences favoring one particular choice alternative. They reported that dissensus of opinion prior to discussion gives the information that is pooled during discussion the greatest opportunity to influence decision making. They note that further research is needed to better understand the situational variations within which the pooling of information during discussion is most likely to impact the decisions that groups eventually make.

Michaelsen, Watson, and Black (1998) examined the quality of individual versus group decision making in a situation that was representative of everyday work situations. They found that a vast majority of groups can outperform their most knowledgeable member on decision-making tasks. They suggest that an awareness of

group processes can have a positive effect on group decision-making effectiveness (p. 9).

Purpose of the Present Study

Given the effects of group processes on the decisions by groups, it is no surprise that the different standard setting procedures produce different passing scores. Rather than trying to find a definitive standard setting methodology, the focus of the standard setting studies should concentrate on the interaction of group members, the instructions given to the group, or the individual differences of group members. Since decision makers have difficulty accessing the information that discussion is presumed to elicit (Larsen, Christensen, Abbott, & Franz, 1996), the best standard-setting method may be one that minimizes the effects of the group influences by simplifying the process. It is hoped that by examining the effects of deliberation style on the outcome of standard setting studies that a better understanding of the effects that methods alone have on the results of the standard setting studies.

The primary purpose of the present study was to determine the effect of deliberation style and requirement to reach consensus on the passing score. A second purpose was to evaluate their raters' satisfaction and post-decision acceptance of the results in the three different procedures, the modified Angoff, verdict-driven and evidence-driven deliberation styles.

We used a modified Angoff standard setting method (Angoff, 1971) as a benchmark for the two experimental methods because the tests were operational. In the modified Angoff, panelists are asked to estimate the percentage of candidates who would answer each item correctly. For the verdict-driven and evidence-driven

deliberation styles, we asked individual raters to form a yes/no opinion about whether minimally competent candidates would answer the item correctly. In the verdict-driven deliberation style, raters were allowed to form individual preferences, then discuss until each item until consensus was reached. In the evidence-driven deliberation-style, raters began deliberation by exchanging views and continue discussion to reach a consensus yes/no decision about each item without forming individual opinions.

Method

Examinations

There were two forms of jurisprudence and professional ethics examinations, Forms 1 and 2. Both forms were developed using the same specifications. Each form consisted of 100 multiple-choice items selected from a pool of items covering legal and ethical aspects of practice including psychotherapeutic relationships, interventions and assessments, records, and professional practice issues.

Behavioral Descriptors or Incidents

The behavioral descriptors or incidents (Flanagan, 1954) were developed by a panel of 10 licensed psychologists who generated behaviors that describe different levels of candidate performance: highly effective, minimally competent, and ineffective. Panelists were asked to link their incidents to tasks and knowledge identified in a practice analysis of psychologists in the United States and Canada. Two hundred and ninety-one behaviors were generated and were "spiraled" to create a questionnaire. The spiraling procedure ensured that judgments about each behavior were rated without knowledge of the original content areas or the original levels of performance for which they were developed.

Of the 291 behaviors generated, 134 behaviors were clearly assigned to one of eight content areas (Assessment, Diagnosis, Intervention, Legal Mandates, Professional Ethics, Limitations and Judgments, Human Diversity, Quality Assurance) and levels of performance. The retranslation criterion was defined by a conjunctive rule: A behavior had to be named by at least 50% of the respondents to a specific content area and the same behavior had a standard deviation that did not exceed 2.00 in terms of level of effectiveness. The nine levels of performance were truncated into three general levels of performance according the mean effectiveness rating: ineffective (1.00-3.00), minimally competent (4.00-6.00), and highly effective (7.00-9.00).

The 33 behaviors assigned to Legal Mandates and Professional Ethics were used as minimum competence criteria in this study.

Design and Procedures

Participants

Forty-seven licensed psychologists served as raters. They were assigned to one of six groups (see Table 1):

(a) Group 1: Form 1: Method A (percentage, with prior ratings)

(b) Group 2: Form 1: Method B (yes-no, consensus with prior ratings)

(c) Group 3: Form 1: Method C (yes-no, consensus without prior ratings)

(d) Group 4: Form 2: Method A (percentage, with prior ratings)

(e) Group 5: Form 2: Method B (yes-no, consensus with prior ratings)

(f) Group 6: Form 2: Method C (yes-no, consensus without prior ratings)

Training

Prior to rating operational items, groups were asked to review the criteria developed with the critical incident technique, the eligibility requirements for licensure, and the meaning of minimal competence and entry-level practice in California. Essentially, the raters' task was to review the criteria in terms of minimal competence to establish the passing score. Appendix A presents a sample of the minimal competence criteria.

Raters were allowed to discuss the concept of minimal competence until they were comfortable continuing with the process. After the discussion, raters took the examination under test-like conditions to gain a candidate's perspective of the difficulty of the items. Raters were then provided with the correct answers to each item. Then, raters were trained in the standard setting procedures with 5-10 practice items so that they could learn to apply the minimum competence criteria to the items.

Method A - percentage, with prior ratings. Raters in the percentage with prior ratings group were asked to independently estimate without discussion, the percentage of minimally competent candidates who would answer each item correctly relative to the behavioral descriptors or incidents. Once individual estimates were completed, raters discussed their ratings for each item and were allowed to revise ratings in light of the discussion.

Method B - yes-no, consensus with prior ratings. Raters in the yes-no, consensus with prior ratings group were asked to independently affirm or disaffirm if minimally competent candidates would answer each item correctly relative to the behavioral descriptors or incidents. Once individual estimates were completed, raters

discussed their ratings for each item until consensus was reached. A single yes-no decision was recorded for each item.

Method C - yes-no, consensus without prior ratings. Raters in the yes-no, consensus without prior ratings group were asked to collectively affirm or disaffirm if minimally competent candidates would answer each item correctly relative to the behavioral descriptors or incidents. Raters discussed each item until consensus was reached. A single yes-no decision was recorded for each item.

Computation of passing scores

The passing score for Groups 1 and 4 (percentage, with prior ratings) was determined by summing each rater's ratings and averaging them across raters. The passing score for Groups 2 and 5 (yes-no, consensus with prior ratings) as well as Groups 3 and 6 (yes-no, consensus without prior ratings) was determined by summing the collective yes decisions made by each group.

Postsession questionnaire

Raters were asked to complete a post-session questionnaire that solicited individual responses (strongly disagree-1 to strongly agree 10).

Questions pertaining to acceptance of group decision

1. I was confident with the group decision.
2. The methodology used for making decisions allowed me to use my judgment.
3. I was able to use the minimum competence criteria when making my decision.
4. I felt that I had an opportunity to actively participate in the discussion.

Question pertaining to reactions to the decision procedure

1. I believe that the passing score will be fair.
2. I was able to follow the instructions.
3. The criterion-referenced passing score process made sense as it was presented in the workshop.
4. The decision process was objective.
5. The decision process was fair.
6. I felt at ease during the discussion.

Percentage of time that opinion changed as a function of group discussion. The final question asked raters to indicate the percentage of time that they changed their opinion as a function of group discussion.

Results

Of the three methods, Method C (yes-no, consensus without prior ratings) produced the lowest passing scores, 74 and 72, as compared to Methods A or B. Method B (yes-no, consensus with prior ratings) produced the highest passing scores, 89 and 85, respectively. Method A (percentage, with prior ratings) produced passing scores, 84 and 76, respectively, that were midway between Method B and Method C (see Table 2).

Postsession responses to questions regarding acceptance of the group decision suggested that raters using Methods A, B, and C were satisfied with and confident about the group decisions (see Tables 3 and 4). Some raters who used Methods A and B were less satisfied or confident than raters who used Method C. Although the pattern of responses was unclear, deliberation style appears to interact with raters' reactions to

decision procedures (see Tables 5 and 6). The lowest mean rating, 6.89 out of a possible 10.00, was reported by raters in Group 4 (Form 2, Method A) to the statement "I felt at ease during the session" whereas the highest mean rating, 10.00 out of a possible 10.00, was reported by raters in Group 3 (Form 1, Method C) "I was able to follow the instructions."

Furthermore, raters who used Methods A (percentage, prior ratings) and B (yes-no, consensus with prior ratings) indicated that they sometimes changed their opinion (11.33%, 18.33%) or frequently changed their opinion (22.50%, 28.13%) after listening to group discussion of the ratings (see Table 7).

During the debriefing that followed each workshop, some raters in Method A (percentage, with prior ratings) commented that the process was objective and fair. Another judge commented that a fixed percentage should be use as the passing score. Still another judge commented that the majority of the questions did not need to be discussed.

The comments for Method B (yes-no, consensus with prior ratings) were also varied. One judge indicated that the method was at times frustrating and time-consuming. Two raters commented that that the method was an impressive way to establish a passing score. Another judge commented that the process was very thorough.

Several raters trained in Method C (yes-no, consensus without prior ratings) indicated that the method was fun, interesting, effective, easy to use, and worked well yet and one judge commented that it was unclear why this method was better than

selecting a fixed percentage score. One judge suggested that each judge's final decision should be anonymous from the rest of the group.

Discussion

An important finding of the present study is the feasibility of using a simple consensus rule that requires a yes-no response and no prior ratings (Method C) to establish a passing score. Consensus decisions provide opportunities for all raters to express and consider all, not some, of the issues in an examination before making final decisions. The procedures for this decision rule result in a passing score that is perceived as acceptable and fair, and more importantly, easy to understand and implement. The raters were able to arrive at final decisions without changing their mind repeatedly with either Form 1 or Form 2. Moreover, the raters readily accepted the group decision and indicated that the process was objective and fair.

Overall, raters who used Method A (percentage, with prior ratings) established passing scores that were midway between passing scores established by raters who used Method B (yes-no, consensus with prior ratings) and raters who used Method C (yes-no, consensus without prior ratings). The results are consistent with previous work (e. g., Chinn & Hertz, 2002) but contrast with other studies (e.g., Impara and Plake, 1997) that have found similar passing scores for percentage and yes-no standard setting methods.

One explanation is that the deliberation style affected the raters' opportunities to participate in the decision-making process. In Method A, raters could establish their prior ratings on a continuum from 25-95% and consider some or all of the information discussed by the group prior to establishing their final ratings. Since the raters were not

required to reach consensus, they were likely to establish final ratings close to or near their prior ratings. The percentage type rating provided them the mechanism for making decisions that were close or near their prior ratings. Thus, they could maintain their opinions without conceding entirely to minority or majority factions.

In Method B, raters could establish their individual dichotomous (yes-no) ratings and then consider all of the information discussed by the group prior to establishing dichotomous (yes-no) ratings as a group. The raters may have initially favored their prior ratings, but were unable to maintain their ratings unless they conceded to majority factions regarding the yes-no rating, hence, the relatively high passing score. The decision environment in Method B appeared to produce enough interpersonal conflict and prolonged discussion that caused the raters to take the full time period allotted for the workshop.

In Method C, raters simply had to establish final ratings as a group. Raters did not have to defend individual opinions but participate and evaluate the information discussed by the group prior to establishing dichotomous (yes-no) ratings as a group. The decision environment appeared to produce very little conflict between raters who used Method C (Groups 3 and 6) such that the groups established their passing scores two hours earlier than raters who used Methods A or B. In essence, there was less pressure to defend individual opinions and more opportunity for interchange of individual opinions. Thus, the decision-making process begins immediately before majority and minority sectional differences are emphasized (Kameda and Sugimori, 1995) thereby maximizing the effects of informational influences on the decisions to be made (Fitzpatrick, 1989).

A second explanation may be the influences of social, motivational and cognitive factors that influence the group dynamics of the raters who used each method. The sharing of information in the different methods explored in the present study could be serving different functions that are not attributable to simple majority or minority preferences. The exchange of information during discussion may have served several purposes, particularly when the task involves interpretative rather than inferential processes (Clark et al., 2000). One purpose could be to provide opportunities for either problem resolution or information seeking. Groups may have begun the early stages of standard setting spending efforts to seek additional information to clarify the context. The group's strategy, either dialogue or information seeking, will influence decision quality and satisfaction with the group process. A standard setting method, such as Method C, could have encouraged dialogue, and thus, information exchange and depth of information processing. Because individuals in Methods A and B were allowed to establish their individual preferences before group discussion, the standard setting methods could encourage information seeking to the point where it was counterproductive. In particular, Method B, there were limited the opportunities for true dialogue because the final decisions were dichotomous and required agreement from all raters.

The small differences in the postsession responses may have been a result of the consistency of the procedures and in the thoroughness of the training provided to the raters. It appears that regardless of the method, the raters are fairly pleased with the process. A true difference was found in the passing score, most notably in the consensus without prior ratings. The raters were able to complete the standard setting

study in less time, approved of the method, and produced a passing score lower than the Angoff percentage method which is often criticized as producing a passing score unreasonably high.

In sum, the present findings offer some insights into the interaction of group members and the instruction given to group members involved in standard setting. The. Method C (yes-no, consensus without prior ratings) shows promise because it is a simple and intuitive process that is easier to use than the most often used method of Angoff (percentage).

Future research could explore in detail the social aspects of raters' acceptance of the group decision (satisfaction and confidence about their decisions), reactions to the decision procedures (fairness, objectivity of the process, perceptions of group members' participation, ease of use of the decision rule) and the reasons why they changed their opinions after discussion of the issues. The data could be used to gain insights on the group processes that underlie standard setting procedures and outcomes.

References

Angoff, W. H. (1971). Scales, norms, and equivalent scores. In R. L. Thorndike (Ed.), Educational measurement (2nd ed., pp. 508-600). Washington, DC: American Council of Education.

Behuniak, P., Archambault, F. X., & Gable, R. K. (1982). Angoff and Nedelskiy standard setting procedures: Implications for the validity of proficiency test score interpretation. Educational and Psychological Measurement, 9, 215-235.

Chinn, R. N., & Hertz, N. R. (2002). Alternative approaches to standard setting for licensing and certification examinations. Applied Measurement in Education, 15, 1-14.

Clark, M. A., Anand, V., & Roberson, L. (2000). Resolving meaning in terpretation in diverse decision-making groups. Group Dynamics: Theory Research and Practice, 4, 211-221.

Davis, J. H., Holt, R. W., Spitzer, C. E., & Stasser, G. (1981). The effects of consensus requirements and multiple decisions on mock juror verdict preferences. Journal of Experimental Social Psychology, 17, 1-15.

Davis, J.H., Kameda, T., Parks, C., Stasson, M., & Zimmerman, S. (1989). Some social mechanics of group decision making: The distribution of opinion, polling sequence, and implications for change. Journal of Personality and Social Psychology, 57, 1000-1002.

Fehrmann, M. L., Woehr, D. J , & Arthur, W. (1991). The Angoff cutoff score method: the impact of frame-of-reference training. Educational and Psychological Measurement, 51, 857-872.

Fitzpatrick, A. R. (1989). Social influences in standard setting: The effects of social interaction on group judgments. Review of Educational Research, 59, 315-328.

Flanagan, J. C. (1954). The critical incident technique. Psychological Bulletin, 51, 327-358.

Impara, J. C., & Plake, B. S. (1997). Standard setting: An alternative approach. Journal of Educational Measurement, 34, 353-366.

Jaeger, R. M. (1989). Certification of student competence. In R. L. Linn (Ed.), Educational Measurement (3rd ed., pp 485-514). New York: Macmillan.

Kameda, T. (1991). Procedural influences in small-group decision making deliberation style and assigned decision rule. Journal of Personality and Social Psychology, 61, 245-256.

Kameda, T., Ohtsubo, Y., & Takazawa, M. (1997). Centrality in sociocognitive networks and social influence: An illustration in group decision-making context. Journal of Personality and Social Psychology, 73, 296-309.

Kameda, T., & Sugimori, S. (1993). Psychological entrapment in group decision making: An assigned decision rule and a groupthink phenomenon. Journal of Personality and Social Psychology, 65, 282-292.

Kameda, T., & Sugimori, S. (1995). Procedural influence in two-step group decision making: Power of local majorities in consensus formation. Journal of Personality and Social Psychology, 69, 865-876.

Larsen, J. R. Jr., Christensen, C., Abbott, A. S., & Franz, T. M. (1996). Diagnosing groups: Charting the flow of information in medical decision-making teams. Journal of Personality and Social Psychology, 71, 315-330.

Michaelsen, L.K., Watson, W.E., & Black, R.H. (1989) A realistic test of individual versus group consensus decision making. Journal of Applied Psychology, 74, 834-839.

Miller, C. E. (1989). The social psychological effects of group decision rules. In: P. Paulus (Ed.), Psychology of group influence (2nd ed., pp. 327-355). Hillsdale, NJ: Lawrence Erlbaum.

Plake, B. S., Impara, J. C., & Irwin, P. M. (2000). Consistentcy of Angoff-based precisions of item performance: Evidence of technical quality of results from the Angoff standard setting method. Applied Measurement in Education, 37, 347-355.

Postmes, T., Spears, R. & Cihangir, S. (2001). Quality of decision making and group norms. Journal of Personality and Social Psychology, 80, 918-930.

Stasson, M. F., Kameda, T., & Davis, J. H. (1997). Group Dynamics: Theory Research and Practice, 1, 316-323.

Winquist, J.R., & Larson, J.R. (1998). Information pooling when it impacts group decision making. Journal of Personality and Social Psychology, 74, 371-377.

Wittenbaum, G. M., Hubbell, A. P. & Zuckerman, C. (1999). Mutual enhancement: Toward an understanding of the collective preference for shared information. Journal of Personality and Social Psychology, 77(5), 967-978.

Author Note

Correspondence concerning this article should be addressed to Norman R. Hertz, Ph.D., HZ Assessments, 177 Arbuckle Avenue, Folsom, California, 95630. Electronic mail may be sent via Internet to hzmail@attbi.com.

Appendix A

Sample of performance behaviors

HIGHLY EFFECTIVE

- Can apply basic legal mandates to novel situations requiring analysis and application beyond current laws.
- Demonstrates recognition of implications of federal and state statues as applied to professional practice.

EFFECTIVE

- Demonstrates knowledge state statutes regarding professional practice
- Describes effects of mandated breaches of confidentiality on therapeutic relationship or treatment prognosis.
- Complies with legal mandates even when laws conflict with own view of optimal treatment.

INEFFECTIVE

- Confuses privilege and confidentiality.
- Cites legal mandates without noting clinical consequences.
- Confuses mandated reporting with optional reporting.
- Reports child, dependent adult, or elder abuse without evaluating client.

Table 1

Groups of Participants by Method and Examination Form

	Method		
	A	B	C
Form	Percentage, with prior ratings	Yes-No, consensus with prior ratings	Yes-No, consensus without prior ratings
1	Group 1 (N=8)	Group 2 (N=7)	Group 3 (N=7)
2	Group 4 (N=9)	Group 5 (N=8)	Group 6 (N=8)

Table 2

Passing Scores for Three Standard Setting Methods

	Method		
Test Form	A	B	C
1	84	89	74
2	76	85	72

Table 3

Mean Ratings Regarding Acceptance of Group Decision for Form 1

	Method		
	A	B	C
1. I was confident with the group decision.	7.86	8.83	9.71
2. The methodology used for making decisions allowed me to use my judgment.	7.86	9.50	9.43
3. I was able to use the concept of minimum competence when making my decision.	7.86	8.50	9.29
4. I felt that I had an opportunity to actively participate in the discussion.	8.29	9.83	10.00

Table 4

Mean Ratings Regarding Acceptance of Group Decision for Form 2

	Method		
	A	B	C
1. I was confident with the group decision.	8.28	8.00	8.69
2. The methodology used for making decisions allowed me to use my judgment.	8.61	8.50	9.38
3. I was able to use the concept of minimum competence when making my decision.	8.39	8.50	8.50
4. I felt that I had an opportunity to actively participate in the discussion.	9.78	9.63	9.88

Table 5

Mean Ratings Regarding Group Reactions to Decision Procedures for Form 1

	Method		
	A	B	C
1. I believe the passing score will be fair.	8.00	8.83	9.71
2. I was able to follow the instructions.	8.29	9.67	10.00
3. The criterion-referenced passing score process made sense as it was presented in the workshop.	7.71	9.17	9.64
4. The decision process was objective.	7.57	8.17	9.07
5. The decision process was fair.	8.14	9.33	9.71
6. I felt at ease during the session.	8.00	8.17	9.43

Table 6

Mean Ratings Regarding Group Reactions to Decision Procedures for Form 2

	Method		
	A	B	C
1. I believe the passing score will be fair.	8.39	7.25	8.69
2. I was able to follow the instructions.	9.22	9.43	9.50
3. The criterion-referenced passing score process made sense as it was presented in the workshop.	9.11	8.00	9.00
4. The decision process was objective.	8.78	7.75	8.88
5. The decision process was fair.	9.00	8.00	9.25
6. I felt at ease during the session.	6.89	8.38	7.69

Table 7

Reported Change of Opinion During Discussion

	Method		
Test Form	A	B	C
1	11.33%	18.33%	15.71%
2	22.50%	28.13%	18.38%